エントロピーを用いた暗号化通信識別方式の検討

Investigation of Identifying Encrypted Communication by Entropy Measurement

神田 敦・システム分科会・情報セキュリティ大学院大学（秋入学）

Abstract: Due to the recent trend of encryption and multiplexing, there's a concern that the conventional network intrusion detection techniques like signature matching will be outdated in the near future. Our research aims to develop an anomaly detection technology for the era of encrypted communication. As a first step of our research, we focused on SSL/TLS, widely used encryption technology, and entropy, well-known indicator for encryption, to examine the ability of identifying a traffic.

問題意識

【想定脅威】
攻撃者が非正規ポートを使って暗号化通信
（C&C: Command & Controlなど）

【課題】
どのポートでどの暗号化プロトコルを使うか不明
→ 全ポートでプロトコル解析するのは非現実的
（処理負荷、プライバシー）

着目技術：エントロピー

シャノンエントロピー（平均情報量）H(X)
情報理論の分野において「情報の不確かさ」を表す

$$H(X) = -\sum_i P_i \log_2 P_i$$

- ✓ 暗号化/パッキング箇所の特定に利用
  （暗号化/パッキングデータは高エントロピー ）
- ✓ プロトコル非依存
- ✓ シンプルな計算

リサーチクエスチョン

【エントロピー を用いた既存研究の課題】
対象とするプロトコルや暗号スイートを限定

【リサーチクエスチョン】
暗号スイートやプロトコルの違いに関わらず
利用可能なエントロピー 特性は？

実験

【データセット】
Alexaランク上位のHTTPS（TLS1.3）対応サイト
へのアクセストラフィック（pcap）
全12,551セッション

【エントロピーの計算方法】

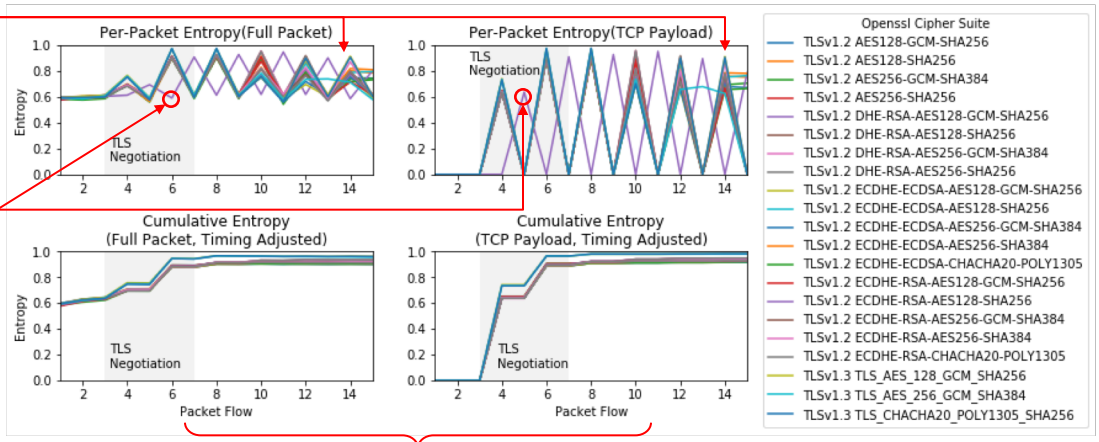
・パケット単位 ・セッション累積	×	・パケット全体 ・TCPペイロード

【分析結果】
- ➢ パケット単体エントロピーは特徴量には不向き
- ➢ パケット全体よりもTCPペイロードの方が
  特徴がより顕著に現れるため有用
- ➢ 「エントロピーの立上り」が特徴量の有力候補

制御パケット(ACK)による値のばたつき（パケット単位）

再送による位相のずれ（パケット単位）

TCPペイロードの方が値の変動幅が大きい（パケット単体、セッション累積）



図：同一ウェブサイトアクセス時のエントロピー推移

今後

- ➢ 追加実験（平文、他プロトコルとの比較）を踏まえた異常検知ロジックの構築
- ➢ 異常検知性能の検証